Vol. XLVI. DECEMBER, 1901.

AMERICAN SOCIETY OF CIVIL ENGINEERS.

INSTITUTED 1852.

# TRANSACTIONS.

This Society is not responsible, as a body, for the facts and opinions advanced in any of its publications.

No. 905.

## SHORT AND EASY METHODS FOR COMPUTING PROBABLE ERRORS.

By E. A. Fuertes, M. Am. Soc. C. E.

The engineer is constantly engaged in a large number of professional problems, the solutions of which depend upon the averages of experiments based upon a certain number of repeated observations, as, for example, when gauging the delivery of a stream by taking many velocity measurements; making a series of tests on cement briquettes; or experimenting with the strength of materials, etc.

Usually, the means of various results are classified, in some more or less arbitrary manner, and these means are accepted as the best obtainable value; but the engineer is not able to determine the actual degree of confidence to be placed in the average value thus obtained; neither can he determine whether the relation between the individual results and their means are the proper ones, nor that the number of observations and experiments are sufficient to justify the blind adoption of their means.

Generally, it has been deemed necessary to apply the method of least squares to discuss the results of important experiments. This method is generally long and tedious, and many engineers are not familiar with it, even if they ever learned it at school. Also, though

the advantage of the rigorous method of least squares is easily granted, often its application would be entirely useless, because the labor and time required would seem unjustifiable in a very large number of cases in engineering investigations; and yet it would be always desirable to learn if the various observations are consistent with each other, and if their mean is entitled to the proper degree of confidence.

If this useful information could be obtained with ease and rapidity, there is no doubt that the engineer would avail himself of its advantages; and the absurd and even dangerous deductions, sometimes based upon insecure averages, would be less numerous, thus fostering reliable progress in a large number of engineering investigations.

The origin of the method herein explained has an interesting history, connected with our Civil War in an indirect way.

When Napoleon III conceived the Quixotic notion of re-establishing the supremacy of the Latin races over our Western Hemisphere he took advantage of the disordered condition of the United States, during our Civil War, to enlist against us the hatred and fear of Spain with regard to her then West Indies possessions, and spurred the commercial jealousies of England, to form a triple alliance between France, England and Spain, aiming at the destruction of our national supremacy.

During the darkest days of the Civil War, and when we were unable to prevent it, the allied armies landed in Mexico, to found a new monarchy upon our southern frontier.

For reasons rather foreign to the objects of this paper, Spain and England withdrew their forces from American soil, leaving Napoleon in the lurch. The Emperor Maximilian had not yet lost the prestige of that formidable alliance, when the Mexican armies had taken to the woods, initiating the guerilla warfare which culminated in the capture of the Emperor, the collapse of the Empire, the pathetic madness of the Empress Carlotta, and the death of the misguided and betrayed Emperor. The Mexican armies operated in the fastness of virgin, tropical forests, and their movements could be guided only by the astronomical work of F. D. Covarrubias, a Mexican engineer of great ability and still greater originality.

The sextant used by him for the purposes of directing the peregrinations of the Mexican forces became injured beyond repair by the fall.

over a precipice, of the mule carrying his field equipment. Mr. Covarrubias, being unable to find the position of his army by any astronomical process involving the knowledge of angular measures, set to work on the spot to develop new methods of computation with entire independence of the graduation of the sextant, using only chronometer times and the almanac data for the solution of astronomical triangles. He felt obliged to modify many astronomical formulas, including those necessary for the discussion of results by least squares. These latter formulas he made exceedingly simple of application, and yet capable of furnishing all the necessary accuracy for a large number of cases in engineering practice.

There are several least-square formulas for various purposes, generally known by the names of their inventors; but they all demand considerable labor on the part of the computer. The Covarrubias formulas can be reduced to a table demanding only a few seconds of time to learn:

*First.*—The probable error of each individual result in any series of experiments;

*Second.*—The probable error of the mean of all the results;

*Third.*—They furnish a fair criterion for the number of repetitions required to reduce the error of the mean to any desirable limit of accuracy;

*Fourth.*—This method avoids the necessity of squaring each one of the residuals, or the squares of the difference between each result and the mean of all the results;

*Fifth.*—It avoids the necessity of taking the algebraic sum of the residuals and nearly all the long and tedious processes of least squares;

*Sixth.*—Any pocket table of logarithms can be used to find the probable errors, without the need of interpolations;

*Seventh.*—The formulas can be reduced to a table, using as a single argument the number of observations taken.

The formulas chosen by Covarrubias are:

$$r = \pm 0.8453 \frac{[V]}{\sqrt{n(n-1)}} = F[V] \quad \ldots\ldots (1)^*$$

$$r_o = \frac{r}{\sqrt{n}} = fr \quad \ldots\ldots (2)$$

* See Doolittle's Practical Astronomy, p. 21. It may be often desirable to use: $r = \frac{0.85\,[V]}{n}$.

In these formulas:

$r$ = the probable error of each single result;

$r_o$ = the probable error of the mean of all the observations or results;

$V$ = the residuals, or the numbers obtained by subtracting each result from the mean of all of them (without regard to signs);

$[V]$ = the arithmetical sum of the residuals;

$n$ = the number of observations taken.

The values of $F$ and $f$, reduced to a table for any values of $n$, are given in Table No. 1.

TABLE No. 1.—LOGARITHMS OF $F$ AND $f$ FOR COMPUTING PROBABLE ERRORS.

| $n$. | Log. $F$. | Log. $f$. | $n$. | Log. $F$. | Log. $f$. |
|---|---|---|---|---|---|
| 2 | 9.7765 | 9.8495 | 26 | 8.5206 | 9.2925 |
| 3 | 9.5379 | 9.7614 | 27 | 8.5038 | 9.2843 |
| 4 | 9.3874 | 9.6990 | 28 | 8.4877 | 9.2764 |
| 5 | 9.2765 | 9.6505 | 29 | 8.4722 | 9.2688 |
| 6 | 9.1884 | 9.6109 | 30 | 8.4572 | 9.2614 |
| 7 | 9.1154 | 9.5774 | 31 | 8.4428 | 9.2543 |
| 8 | 9.0529 | 9.5485 | 32 | 8.4288 | 9.2474 |
| 9 | 8.9983 | 9.5229 | 33 | 8.4152 | 9.2407 |
| 10 | 8.9499 | 9.5000 | 34 | 8.4020 | 9.2343 |
| 11 | 8.9063 | 9.4793 | 35 | 8.3892 | 9.2280 |
| 12 | 8.8667 | 9.4604 | 36 | 8.3768 | 9.2218 |
| 13 | 8.8304 | 9.4430 | 37 | 8.3648 | 9.2159 |
| 14 | 8.7970 | 9.4269 | 38 | 8.3530 | 9.2101 |
| 15 | 8.7659 | 9.4119 | 39 | 8.3416 | 9.2045 |
| 16 | 8.7369 | 9.3979 | 40 | 8.3304 | 9.1990 |
| 17 | 8.7097 | 9.3848 | 41 | 8.3196 | 9.1936 |
| 18 | 8.6841 | 9.3724 | 42 | 8.3090 | 9.1884 |
| 19 | 8.6600 | 9.3606 | 43 | 8.2987 | 9.1833 |
| 20 | 8.6371 | 9.3495 | 44 | 8.2886 | 9.1783 |
| 21 | 8.6154 | 9.3389 | 45 | 8.2787 | 9.1734 |
| 22 | 8.5947 | 9.3288 | 46 | 8.2690 | 9.1686 |
| 23 | 8.5749 | 9.3191 | 47 | 8.2596 | 9.1639 |
| 24 | 8.5560 | 9.3099 | 48 | 8.2503 | 9.1594 |
| 25 | 8.5379 | 9.3010 | 49 | 8.2413 | 9.1549 |
| | | | 50 | 8.2324 | 9.1505 |

*Example I.*—The following five observations were taken to find the error of a clock:

| Observation. | Clock fast. | $V$ = residuals. | |
|---|---|---|---|
| | Min. Sec. | Sec. | The residuals $V$ are obtained by finding the difference between each clock error and the mean of all the errors. |
| I | 16 41.72 | 0.18 | |
| II | 16 41.18 | 0.36 | |
| III | 16 41.15 | 0.39 | |
| IV | 16 41.60 | 0.06 | |
| V | 16 42.06 | 0.52 | $[V]$ = the arithmetical sum of the residuals. |
| | 16 41.54 = mean | $[V] = 1.51$ = sum | |

Since, in this example, there were five observations, we take out the values of $F$ and $f$ corresponding to $n = 5$, proceeding as follows:

Log. $F$..... 9.2765 from the table,
Log. $[V] = (1.51)$ 0.1790 = log. of sum of residuals,
Sum..... 9.4555 = log. $r$: hence, $r = \pm 0.28$ second,
Log. $f$..... 9.6505 from the table,
Sum..... 9.1060 = log. $r_o$: hence, $r_o \pm 0.13$ second.

This shows that the probable error of each single determination, $= r$, is within 0.28 second, more or less; and the probable error of the mean of all observations, $= r_o$, can be depended upon to be true within $\pm 0.13$ second, that is: considering the number and nature of the observations; the accuracy with which the decimals have been carried out in the original time computations; the outstanding errors of the stars chosen, as well as of the instruments used; accidental errors and every other circumstance affecting the error-producing causes attending all classes of observations, the probable errors, $r$ and $r_o$ measure the degree of reliability of the work.*

The value of $r$ must vary with each individual observer in every case in which the result depends upon personal skill; but it is very apt to remain constant for the same observer after he has acquired a reasonable degree of practice in the special manipulation of the work in hand.

If, in the above example, the observer should require a better degree of precision in the probable error of the mean than 0.13 second, since, from the value found for $r$, he knows that the probable error of each observation is about 0.3 second, if he should now desire to reduce the error of the mean to 0.07 second, he must make

$$r_o = \frac{r}{n} = 0.07 \text{ second},$$

Solving for $n$, we get $n = 16$; because $r_o = \frac{0.3}{16} = 0.07$ second.

Therefore, the observer must take 16 observations, instead of 5, to obtain the desired result.

*Example II.*—The following observations were taken to ascertain the relation between the velocity of water, as derived directly from its actual weight when passing through a known cross-section in a given time, and the velocity as deduced from Pitot tube observations, viz.:

$$\frac{\text{Velocity by weight}}{\text{Velocity by Pitot tube}} = \text{Coefficient}.$$

*The rigorous method of least squares gives $r = \pm 0.26$ second, and $r_o = \pm 0.12$ second.

| Set A. | | Set B. | |
|---|---|---|---|
| Coefficient. | Residuals. | Coefficient. | Residuals. |
| 0.8400 | 0.0055 | 0.8280 | 0.0020 |
| 0.8349 | 0.0004 | 0.8260 | 0.0000 |
| 0.8409 | 0.0064 | 0.8328 | 0.0068 |
| 0.8430 | 0.0095 | 0.8299 | 0.0039 |
| 0.8476 | 0.0031 | 0.8179 | 0.0081 |
| 0.8359 | 0.0014 | 0.8268 | 0.0008 |
| 0.8245 | 0.0100 | 0.8346 | 0.0086 |
| 0.8223 | 0.0122 | 0.8139 | 0.0121 |
| 0.8348 | 0.0003 | 0.8310 | 0.0050 |
| 0.8213 | 0.0132 | 0.8190 | 0.0070 |
| Sum = 8.3452 | [V] = 0.0620 | Sum = 8.2599 | [V] = 0.0543 |
| Mean = 0.8345 | | Mean = 0.8260 | |

Entering the table with $n = 10$, we have, for each set:

| Set A. | Set B. |
|---|---|
| Log. $F$..... 8.9499 from the table, | Log. $F$..... 8.9499 from the table, |
| Log. $[V]$..... 8.9724 | Log. $[V]$..... 8.7348 |
| Log. $r$ = sum..... $\overline{7}.9223$, hence, $r = 0.0083$, | Log. $r$ = sum =... $\overline{7}.6847$, hence, $r = 0.0048$, |
| Log. $f$..... 9.5000 from the table, | Log. $f$ =..... 9.5000 from the table, |
| Log. $r_o$ = sum..... $\overline{7}.4223$, hence, $r_o = 0.0026$. | Log. $r_o$ = sum =.. $\overline{7}.1847$, hence, $r_o = 0.0015$. |

The foregoing is a very instructive result, for, as may be learned by remarks entered in the notebook in which the observations were kept, the first set, $A$, was not so good as Set $B$; but the simple use of the table establishes the fact that the probable error of each single experiment in Set $A$ is about twice as large as those of Set $B$, while the probable errors of the means in each set differ by a very much smaller amount.

If we now use Sets $A$ and $B$ as a single set, we shall get, using $n = 20$:

Log. $F$..... 8.6371 by the table,
Log. $[V] = (0.1169)$..... 9.0678

Log. $r$ ..... 7.7049, hence $r = 0.0051$,
Log. $f$..... 9.3455 by the table,

Log. $r_o$..... 7.0504, hence $r_o = 0.0011$;

from which we learn that the probable error of the single result is smaller than even the average of the previous values for each set; while the probable error of the mean result of the entire sets, $A + B$,

is less than one-half of the value found by the averages of the independent sets. This establishes the usefulness of Set *A* to be considered as tending to give accuracy to the final result, and thus the best value of the coefficient sought is $0.83025 \pm 0.0011$.

No claim is made for originality: The intention of the writer has been only to place an easily handled tool in the hands of members of the Society. Extensive practice has proved the usefulness of this approximate method for the discussion of experiments.